

波浪の極値統計解析

合 田 良 実

1. まえがき

海岸、港湾、および海洋構造物の設計では、まず設計波高を定めなければならない。既存の構造物の設計データが使えないときは、新しく資料を収集して設計波の諸元を選定する。この資料としては、長期間にわたる台風や低気圧などによる高波の観測値あるいは推算値を使い、これを極値統計の手法を用いて解析する。

極値統計解析は波浪の問題ばかりでなく、気象統計や河川の洪水流量の解析に古くから利用されており、手法としても確立されているかのように受け取られている。しかし、実際に作業に取りかかってみるといろいろ不明確な点が多いことに気づく。著者は最近この極値統計の問題についてモンテカルロ法による数値実験を行ない、今まであいまいにされてきた事項についてある程度の解答を得ることができた¹⁾。そこで、本稿では実際に波浪の極値統計資料を解析して設計波を選定する場合を対象として、作業手順を中心に解説することにする。なお詳細については文献¹⁾を参照されたい。

2. 波浪の極値資料の収集と分析

(1) 極値資料の分類

洪水流量や高潮などの場合には比較的長い期間にわたる統計資料を得ることができるので、毎年の最大値を対象として解析する場合が多い。こうしたデータは**毎年最大値資料**と呼ばれる。これに対して波浪データの場合には推算値であっても30～40年程度が最長であり、観測値の場合には10年を超える資料を得ることがむずかしい。このため我が国では、顕著な高波毎にその極大値を抽出し、この時系列について統計解析を行なうのが通例である。水文統計でいえば、「部分的水文資料」の中の「非毎年超過値資料²⁾」に相当する。ここでは極大値の時系列資料を単純に**極大値資料**と呼ぶことにする。

なお、外国の論文の中には定時観測値の全データに対して確率分布関数をあてはめ、これを外挿して設計波高を選定することをいまだに提案しているものを見かける。しかし、この方法は統計学的に誤りであるので採用すべきでない。

極値資料としては、対象とすべき全データを収集している場合と、その一部しか収録できない場合とがある。毎年最大値資料であれば統計期間の全ての年について最大値を求めることが容易である。しかし、高波の極大値資料の場合には毎年の高波の発生回数が多いため、全データを取り扱うのが困難であり、あるレベル以上のデータに限定して解析するのが普通である。英語では、前者を“uncensored data”，後者を“censored data”と呼んで区別している。ここでは前者を**全数極値資料**，後者を

部分極値資料と呼び分けておく。本稿で紹介する方法を用いると、この2種類の極値資料ともに共通の方式で解析することが可能である。

(2) 高波の極値データの収集

現地観測資料の場合には、波浪観測台帳について欠測期間の有無を調査する。高波のピークが部分的に欠けているとき、ピーク値がなんらかの方法で補足できなければ、その高波の観測期間は欠測とすべきである。そうして、欠測期間を除外した有効統計年数を計算する。欠測期間が季節的にむらがある場合には、高波の季節特性を考慮して有効統計年数を求める³⁾。この統計期間を K 年とする。波浪推算資料の場合には、対象とする統計期間はあらかじめ設定しているので明らかである。ただし、推算を行なった以外の台風や低気圧によって大きな波浪が発生していないことを確認しておかなければならない。

いずれの場合にも、統計期間中に発生したと思われる高波の全個数を推定し、これを N_T と表しておく。観測・推算資料とも、波高の小さいデータは拾い洩れがあるので、これ以上は確実に拾い出されていると思われる限界波高以上のものを対象とし、それに達しないデータは棄却する。こうして抽出できたデータの個数を N とする。これらのデータから、平均発生率 λ およびデータ採択率 ν を次のように定義する。

$$\lambda = N_T / K, \quad \nu = N / N_T \quad (1)$$

なお、高波の発生総個数 N_T あるいは平均発生率 λ は概略の値でよい。

以上の作業は高波の発生原因、すなわち台風、低気圧、冬型気圧配置、不連続前線などの種類別に分けて行なうことが望ましい。もっとも、観測資料の場合には天気図との照合が必要であり、実行困難な場合もある。

(3) データの並べ替え

全数極値資料でも部分極値資料でも、採択されたデータは大きさの順に並べ替え、順番をつける。波浪データの場合には最大値を1番とするのが普通であり、ここでもその方法を採用する。したがって、

$$x_{(1)} > x_{(2)} > \cdots > x_{(m)} > \cdots > x_{(N)}$$

ただし、 x は極値データであり、波浪資料であれば高波のピーク値である。このように順位付けのされたデータは順序統計量といわれる。

3. プロットイング・ポジションの計算

(1) プロットイング・ポジションの考え方

極値データはいずれもその出現確率が低いものであるが、それぞれどのくらいの確率であるかを決めておかないと、統計解析を行なうことができない。初期の極値統計解析では確率統計紙にデータをプロットしたので、どの位置にプロットするかという意味でプロットイング・ポジションの問題とい

われている。この決めかたについては幾つかの考え方があ。これについて検討するため極値分布の一つである Fisher-Tippett I 型分布 (Gumbel が詳しく解説した⁴⁾ のでグンベル分布ともいわれる) について簡単な数値実験を行なった。すなわち、FT-I 型の母集団から一様乱数を使って 10 個ずつのデータを多数組抽出し、各組について大きさの順に並べ替えて最大値 $x_{(1)}$ を求めた。そして、 $x_{(1)}$ およびその非超過確率 F_1 の頻度分布を求めた結果が図-1 である。サンプル組数は 100 組である。

左下の図は、極値データ $x_{(1)}$ とその非超過確率 F_1 の関係であり、これは当然に 1 本の曲線上にある。左上は F_1 の頻度分布、右横は $x_{(1)}$ の頻度分布である。この例では、最大値 $x_{(1)}$ が 0.84

~6.04 の範囲に広く分布しており、サンプル毎の変動が大きいことが分かる。実際の問題ではただ一つ得られたサンプルにおける順序統計量に対して何らかの非超過確率を割り付けなければならない。この考え方としては、母集団における $x_{(m)}$ の平均 $E[x_{(m)}]$ に対応する確率 $F\{E[x_{(m)}]\}$ を用いるのが妥当である^{5),6)}。したがって、非超過確率の計算式は極値分布関数毎に異なることになる。

(2) プロット公式の具体例

i) ワイブル公式

$$\hat{F}_m = 1 - m / (N_T + 1) \quad (2)$$

ここに、 m は降順の順位、 N_T は対象とする期間中に発生したと考えらるデータの総個数である。

ii) ブロム公式

$$\hat{F}_m = 1 - (m - 3/8) / (N_T + 1/4) \quad (3)$$

iii) グリソルテン公式

$$\hat{F}_m = 1 - (m - 0.44) / (N_T + 0.12) \quad (4)$$

iv) ペトルアスカス・アーガード (P & A) 公式

$$\hat{F}_m = 1 - (m - \alpha') / (N_T + \beta) \quad (5)$$

ここに、

$$\left. \begin{aligned} \alpha' &= 0.30 + 0.18 / k \\ \beta &= 0.21 + 0.32 / k \end{aligned} \right\} \quad (6)$$

なお、 k は後述のワイブル分布の形状母数である。

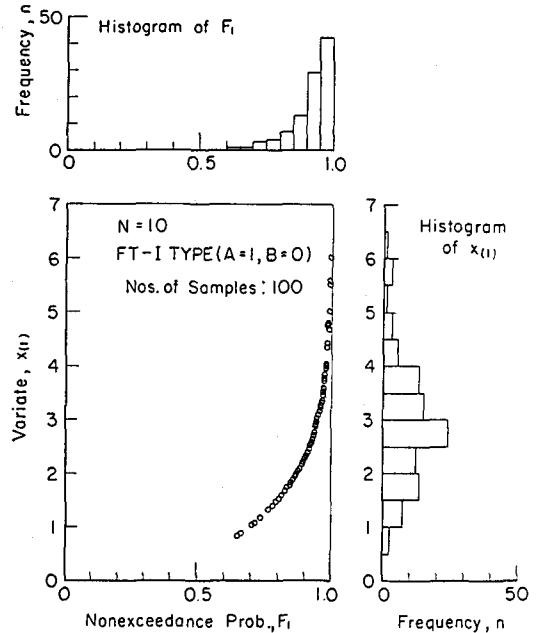


図-1 FT-I 型分布からの極値データのサンプル中の最大値の分布例

v) 同上の修正式

$$\left. \begin{aligned} \alpha' &= 0.20 + 0.27/\sqrt{k} \\ \beta &= 0.20 + 0.23/\sqrt{k} \end{aligned} \right\} \quad (7)$$

以上の諸公式のうち、第1のワイブル公式はグンベルが推奨したこともあって極値分布の関数形を問わずに広く使われている。しかし、Cunnane⁷⁾が厳しく批判したように、この公式は $x_{(m)}$ の平均値に対する非超過確率ではなくて F_m の平均値に基づいており、またこの公式を使うと確率波高に大きな偏り (Bias) をもたらす。したがって、この公式は使用すべきでない。

第2のブロム公式は、Blom⁸⁾が正規分布に対して導いたものであり、対数正規分布に対しても適用すべきものである。その次のグリングオルテン公式は、Gringorten⁹⁾がFT-I型分布に対して求めたものであり、第4のP & A公式はPetruaskas & Aagaard¹⁰⁾がワイブル分布における $x_{(m)}$ の平均値の数値計算結果を公式化したものである。最後のP & A公式に対する修正は著者が行なったもので¹⁾で、原式のままでは確率波高の推定値に負の偏りが出る傾向があるのを補正するためのものである。

プロットング公式にはこの他にもいろいろあるが、波浪の極値統計解析用としては以上で十分であろう。

(3) 数値実験によるプロットング公式の比較

図-2は、FT-I型極値分布について確率波高の推定値が母関数の値からずれている程度を調べた結果である。FT-I型分布で母数を $A=1$ 、 $B=5$ と設定し、一つのサンプル中のデータ個数をあらかじめ定め、一様乱数を使って10,000組のサンプルを抽出した。抽出されたデータは全て使用したので、全数極値資料に当たる。各サンプルに対してFT-I型分布をあてはめて母数を推定し、再現期間がデータ長の10倍の場合の確率波高を算定した。そして、この確率波高の推定値の平均値を求め、これと母関数における確率波高の値との差を百分率で表している。図中の記号は、分布関数のあてはめに用いたプロットング公式に対応しており、ワイブル公式を使うと確率波高が大きめに推定されることが明らかである。推定値の偏り誤差は、サンプルが小さい、すなわちサンプル中のデータ個数が少ないほど大きくなる。なお、Barnett および New と記したのはそれぞれ別の公式を用いた結果であるが、ここでは説明を省略する。

図-3は、ワイブル分布のうち形状母数が $k=1$ の

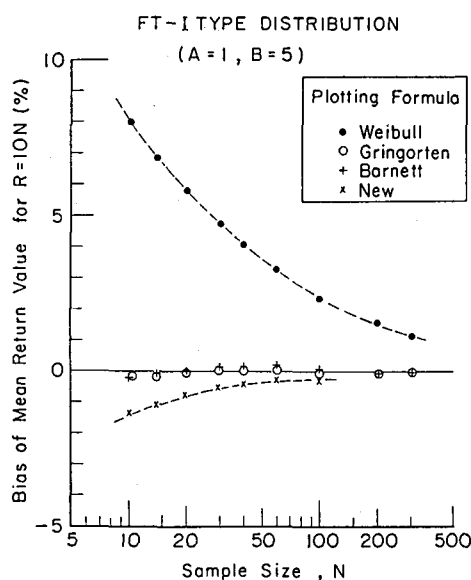


図-2 プロットング公式による確率波高の推定値の偏り量の変化 (FT-I型分布)

場合（これは指数分布に等しい）について同様に調べた結果である。ワイブルのプロットイング公式が大きな偏りをもたらすこと、P & A 公式はやや負の偏りを生むのに対し、修正 P & A 公式ではこの偏り量が減少していることなどが分かる。図の右側は、確率波高の推定値の Root-Mean-Square 誤差（偏りによる誤差を含む）であり、これについてもワイブル公式が不具合であることが明瞭である。

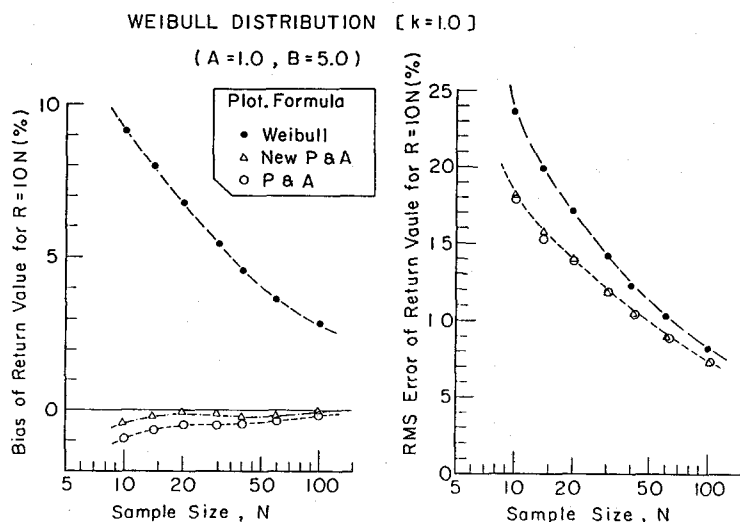


図-3 プロットイング公式による確率波高の推定値の偏り量および RMS 誤差の変化 ($k=1.0$ のワイブル分布)

(4) 部分極値資料に対するプロットイング公式

式(2)~(5)ではデータ個数として従来から使われてきた採択個数 N ではなく、発生総個数 N_T を用いている。これは Muir & Shaarawi⁵⁾ の提案に基づく。二つの考え方の差を図-4で説明する。図の上半は極値データの頻度分布を確率密度関数の形で表したもので、全データのうち x_c を超えるもの(斜線部分)を部分極値資料として採択した場合を考える。プロットイング・ポジションを N_T を使って計算することは、図中の破線で示した限界値以下のデータの存在を考慮していることになる。これに対して、データ個数として N を使うことは限界値以下のデータの存在を無視しているわけで、図の上半の分布を図の下半のように歪めて考えていることになる。このため、与えられたサンプルから元の分布関数の形状を復元することが困難になる。従来のように入手した極値データの個数 N を使う方式は簡単であるが、こうした問題があるので避けるべきである。

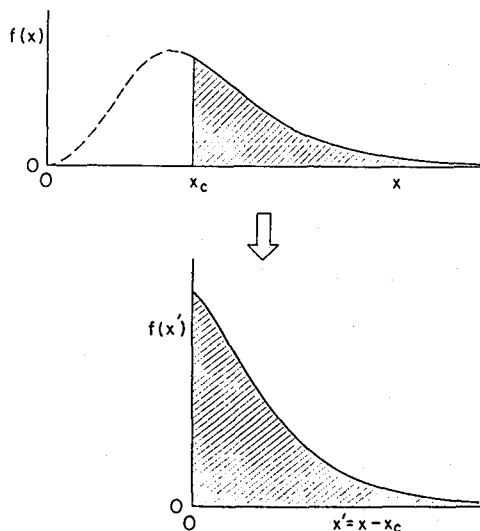


図-4 部分極値資料における確率密度関数の形状

4. 分布関数のあてはめ

(1) 対象とする極値分布関数

極値統計における大きな問題は、対象とする現象の極値分布がどのようなものであるかが分からない

い場合が多いことである。経験的あるいは先験的に特定の分布関数を取り上げ、入手した極値データにその関数をあてはめることがしばしば行われるけれども、その分布関数が適切であったかどうかの判定はむずかしい。自然現象、特に波浪の場合には極値統計解析を行なうために必要な十分に長い期間のデータがほとんど得られていない。このため波浪の極値統計解析としては、複数の極値分布関数を候補として仮定し、データに最も適合する分布を採択するのが合理的と考えられる。

対象とする極値分布関数としては、次の FT - I 型分布ならびにワイブル分布のうちの $k=0.75, 1.0, 1.4$, および 2.0 の 5 種類の中から選択するのがよい。

i) FT - I (Fisher-Tippett type I) 型分布

$$F(x) = \exp \{ - \exp [- (x - B) / A] \} \quad (8)$$

ここに、 A は尺度母数 (scale parameter), B は位置母数 (location parameter) と呼ばれる。いずれも極値データ x と同じ次元を持つ。この分布に従う統計量の母集団平均および標準偏差は次のように与えられる。

$$\mu = E[x] = B + A\gamma \quad (9)$$

$$\sigma = \{ E[x^2] - E^2[x] \}^{1/2} = \pi A / \sqrt{6} \quad (10)$$

ただし、 γ はオイラー定数 ($0.5772\cdots$) である。

ii) ワイブル分布

$$F(x) = 1 - \exp \{ - [(x - B) / A]^k \} \quad (11)$$

ここに、 A, B, k はそれぞれ尺度母数, 位置母数, および形状母数である。また、母集団平均および標準偏差は次のとおりである。

$$\mu = B + A\Gamma(1 + 1/k) \quad (12)$$

$$\sigma = A [\Gamma(1 + 2/k) - \Gamma^2(1 + 1/k)]^{1/2} \quad (13)$$

ただし、 $\Gamma(\quad)$ はガンマ関数である。

(2) 母分布関数への適合度

以上の対象関数の選択は、Petruakas & Aagaard¹⁰⁾に準じたものであるが、原提案よりもワイブル分布の種類を減らしてある。その理由は、データ個数が数十程度の極値資料では本来の母分布関数を復元するのが非常にむずかしいためである。このことを例示したのが図-5である。図の最上段は、FT - I 型極値分布の母集団からデータ個数が10あるいは100のサンプルを一様乱数を使って各10,000組抽出し、各組に最も適合する分布関数を探してそれらの出現率を算出した結果である。いずれも、全数極値資料としての解析である。

あてはめの候補としては、ワイブル分布の $k=0.75, 0.85, 1.0, 1.1, 1.25, 1.5$, および 2.0 , ならびに FT - I 型分布と対数正規分布を取り上げた。最適の判定は、次項に述べる相関係数最大方式によっている。図の横軸は最適関数を記号で表したもので、L - N は対数正規分布を示す。この結果では、母関数が FT - I 型であるにもかかわらず、データ個数が $N=10$ と少ない場合など、

与えられたサンプルが対数正規分布あるいは $k=2.0$ のワイブル分布に最もよく適合すると判断される率が非常に高い。データ個数が100と多くなると母関数である FT-I 型分布を最適とする割合が増えるけれども、それでも全体の34%に過ぎない。

図の中段は $k=2.0$ のワイブル分布を母集団とする場合、下段はおなじく $k=1.0$ の場合について、FT-I 型分布の場合と同様の作業を行なった結果である。 $k=2.0$ のワイブル分布では、データ個数が多くなると母関数に適合する割合が大きいけれども、データ個数が10程度では対数正規分布を最適とするサンプルのほうが多くなる。 $k=1.0$ の場合はさらに母関数の適合率が低い。

このように母分布関数への適合率が低いのは、基本的にはサンプル毎のデータの変動性によるものであり、あてはめ手法あるいは最適の判定方法によるものではないと考えられる。ここでは全数極値資料について示したが、部分極値資料の場合には母関数への適合率が一層低くなる。このため、分布関数のあてはめ候補を多くしても正しい答えが得られる

とは期待できない。そこで、ワイブル分布については $k=0.75, 1.0, 1.4$, および 2.0 の4種類に限定した。また、対数正規分布はその特性が $k=2.0$ のワイブル分布とよく似ていて代替可能である。さらに、対数正規分布を残しておくとも FT-I 型分布に属するサンプルが誤って対数正規分布型と判定されるおそれ大きい。したがって、最初から対数正規分布に属すると判断される場合以外は、これを除外するのが妥当と考えられる。

(3) 分布関数のあてはめの作業手順

ここでは K 年間における高波の極値統計資料が N データ収集されたものとして、作業手順を順番に説明する。なお、この高波の極値データは気象原因別に別個に取りまとめられ、あらかじめ設定した限界値以上のデータは落ちこぼれなく収録されているものとする。

- 1) この期間の高波の総発生個数 N_T の概略値を推定し、平均発生率 λ およびデータ採択率 ν を式(1)で算定する。
- 2) データを並べ替え、最大のものを $H_{(1)}$ 、最小のものを $H_{(N)}$ として順位付けする。
- 3) 波高の順序統計量に対して FT-I 型分布および $k=0.75 \sim 2.0$ のワイブル分布を想定し、そ

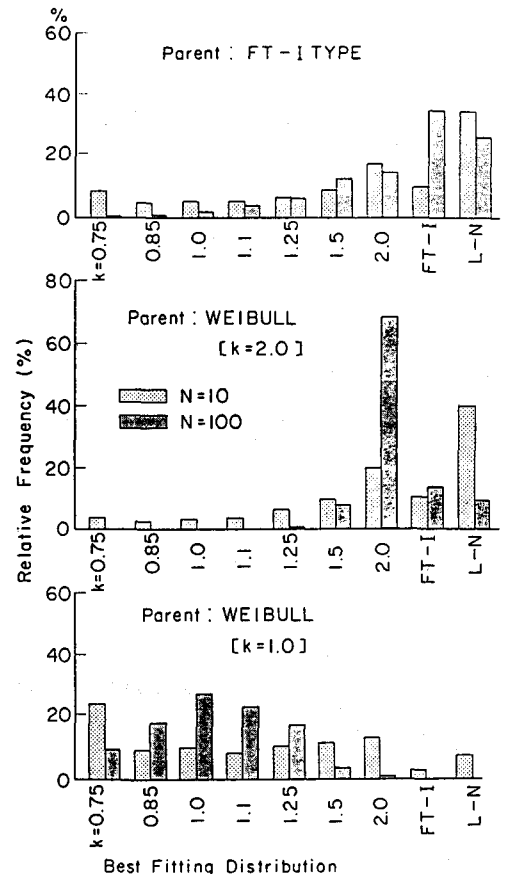


図-5 極値データのサンプルに対する各種分布関数の適合率

れぞれの非超過確率 $\hat{F}_m$ をグリーンゴルテン公式 (式. 4) および修正 P & A 公式 (式. 5 & 7) で計算する。

4) 非超過確率 $\hat{F}_m$ に対し、次式で基準化変量 $y_{(m)}$ を計算する。

FT - I 型分布：

$$y_{(m)} = -\ln [-\ln \hat{F}_m] \quad (14)$$

ワイブル分布：

$$y_{(m)} = [-\ln (1 - \hat{F}_m)]^{1/k} \quad (15)$$

5) 極値波高データ $H_{(m)}$ と基準化変量 $y_{(m)}$ との間に次の直線関係を仮定し、最小 2 乗法を適用して母数推定値 $\hat{A}$ および $\hat{B}$ を求める。

$$H_{(m)} = \hat{A}y_{(m)} + \hat{B} \quad (16)$$

6) 最小 2 乗法を適用するときに $H_{(m)}$ と $y_{(m)}$ との相関係数も同時に求め、FT - I 型分布および 4 種類のワイブル分布をあてはめた中で相関関数が最大のものを最適合の分布関数と判定する。

表-1 波高の極値データに対する非超過確率と基準化変量の計算例

順位 m	極 値 $H(m)$	FT-I 型		ワイブル($k=0.75$)		ワイブル($k=1.0$)		ワイブル($k=1.4$)		ワイブル($k=2.0$)	
		$\hat{F}_m$	$y_{(m)}$	$\hat{F}_m$	$y_{(m)}$	$\hat{F}_m$	$y_{(m)}$	$\hat{F}_m$	$y_{(m)}$	$\hat{F}_m$	$y_{(m)}$
1	8.36	0.9895	4.55	0.9909	7.86	0.9901	4.61	0.9893	2.95	0.9886	2.12
2	7.02	0.9706	3.51	0.9722	5.48	0.9714	3.55	0.9706	2.46	0.9699	1.87
3	6.94	0.9518	3.01	0.9535	4.46	0.9527	3.05	0.9518	2.21	0.9511	1.74
4	6.85	0.9330	2.67	0.9348	3.82	0.9339	2.72	0.9331	2.04	0.9324	1.64
5	6.74	0.9142	2.41	0.9161	3.35	0.9152	2.47	0.9144	1.90	0.9136	1.57
6	6.20	0.8953	2.20	0.8974	3.00	0.8965	2.27	0.8957	1.79	0.8945	1.50
7	5.92	0.8765	2.03	0.8787	2.71	0.8778	2.10	0.8769	1.70	0.8762	1.45
8	5.68	0.8577	1.87	0.8599	2.46	0.8591	1.96	0.8582	1.61	0.8574	1.40
9	5.57	0.8389	1.74	0.8412	2.26	0.8404	1.84	0.8395	1.54	0.8387	1.35
10	5.42	0.8200	1.62	0.8225	2.08	0.8216	1.72	0.8207	1.47	0.8199	1.31
11	5.34	0.8012	1.51	0.8038	1.92	0.8029	1.62	0.8020	1.41	0.8012	1.27
12	5.10	0.7824	1.41	0.7851	1.78	0.7842	1.53	0.7833	1.35	0.7825	1.24
13	5.09	0.7636	1.31	0.7664	1.65	0.7655	1.45	0.7646	1.30	0.7637	1.20
14	4.95	0.7447	1.22	0.7477	1.53	0.7468	1.37	0.7458	1.25	0.7450	1.17
15	4.81	0.7259	1.14	0.7290	1.43	0.7281	1.30	0.7271	1.21	0.7262	1.14
16	4.77	0.7071	1.06	0.7103	1.33	0.7093	1.24	0.7084	1.16	0.7075	1.11
17	4.63	0.6883	0.99	0.6916	1.24	0.6906	1.17	0.6896	1.12	0.6888	1.08
18	4.61	0.6694	0.91	0.6729	1.16	0.6719	1.11	0.6709	1.08	0.6700	1.05
19	4.41	0.6506	0.84	0.6542	1.08	0.6532	1.06	0.6522	1.04	0.6513	1.03
20	4.34	0.6318	0.78	0.6355	1.01	0.6345	1.01	0.6335	1.00	0.6325	1.00
21	4.11	0.6130	0.71	0.6168	0.95	0.6158	0.96	0.6147	0.97	0.6138	0.98
$H=5.565$ $\sigma(H)=1.101$		$\hat{A}=1.091$ $\hat{B}=3.617$ $r=0.9842$		$\hat{A}=0.614$ $\hat{B}=4.029$ $r=0.9621$		$\hat{A}=1.147$ $\hat{B}=3.374$ $r=0.9790$		$\hat{A}=2.084$ $\hat{B}=2.334$ $r=0.9878$		$\hat{A}=3.560$ $\hat{B}=0.786$ $r=0.9910$	

上記の作業手順において最小2乗法を採用したのは、これが一番分りやすい方法であり、かつ部分極値資料も全数極値資料と同じように取り扱えるからである。積率法では、部分極値資料を取り扱うのが困難である。なお、最近是最尤(ゆう)法がしばしば取り上げられるが、これは計算が面倒であり、また確率波高の推定値にやや偏りが出る傾向があるようなので採択しなかった。

表一1は、ある港で台風による高波が波高計によって有効統計年数10.74年にわたって取得されていたので、これを極値統計解析した例である。この有効統計期間中に有義波高のピーク値が4.0m以上の高波が全部で21例観測されていた。表の第2欄はこれら21個の極値データを大きさの順に並べたものである。この有効統計期間の間にこの港に来襲したと考えられる台風性の高波は、天気図などから判断して $N_T=53$ 個と推定された。したがって、平均発生率は $\lambda=53/10.74=4.93$ 個/年であり、データ採択率は $\nu=21/53=0.396$ である。

表の第3欄以降は、FT-I型分布および4種類のワイブル分布に対して式(4)、(5)、および(7)を用いて非超過確率 $\hat{F}_m$ を計算し、その結果からさらに基準化変量 $y(m)$ を式(14)、(15)で求めたものである(実際の計算では有効桁数8桁で実行している)。

表の下3行は、各分布関数について極値データ $H(m)$ と基準化変量 $y(m)$ とに対して最小2乗法を適用して尺度母数および位置母数の推定値 $\hat{A}$ および $\hat{B}$ 、ならびに相関係数 r を求めた結果である。この相関係数の値を比較することにより、 $k=2.0$ のワイブル分布がこのデータに最も適合すると判定される。

5. 確率波高の計算

(1) 再現期間と確率波高

極値統計では「再現期間」(return period)が基本概念の一つとして用いられる。再現期間とは、対象とする事象のうちある特定の値を超えるものが平均して1回起きるような時間間隔をいう。極値事象の値を x_R と特定したとき、これに対する再現期間 R は、母集団の分布関数 $F(x)$ が既知であるとして、次式で計算される。

$$R = 1 / \{ \lambda [1 - F(x_R)] \} \quad (17)$$

これは極大値資料に対するものであり、毎年最大値資料に対しては形式的には $\lambda=1$ と置いたものを使用する。

逆に、再現期間 R を指定したときの極値 x_R も重要な数値である。これは、 R 年確率水文学量、 R 年確率波高などと呼ばれる。この値は、式(17)を書き換えた次式を解いて求めればよい。

$$F(x_R) = 1 - 1/(\lambda R) \quad (18)$$

実際の問題では母集団の分布関数は永久に未知であり、与えられたサンプルに対してあてはめた極値分布関数を使って再現期間 R あるいは R 年確率波高などを推定することになる。

(2) 確率波高の計算式

本稿で取り上げている FT - I 型分布およびワイブル分布における確率波高の計算式は次のとおりである。

$$H_R = \hat{A}y_R + \hat{B} \quad (19)$$

ここに、

$$y_R = -\ln \{-\ln [1 - 1/(\lambda R)]\} : \text{FT - I 型分布} \quad (20)$$

$$y_R = [\ln (\lambda R)]^{1/k} : \text{ワイブル分布} \quad (21)$$

表-1 のデータについて再現期間 $R = 2 \sim 200$ 年の範囲で確率波高を計算すると図-6 のようになる。前節の計算で $k=2.0$ のワイブル分布が最適と判定されたけれども、ここでは5種類の分布関数の全てについて図示している。再現期間が有効統計年数に等しい $R=11$ 年ではどの分布関数もほとんど同一の確率波高を与える。しかし、再現期間が長くなるにつれて差が次第に大きくなり、 $k=0.75$ のワイブル分布が最大、 $k=2.0$ のワイブル分布が最小の確率波高の推定値を与える。

(3) 母分布関数が未知であることによる確率波高の偏りの実験値

先に図-2, 3 で示したように、プロットング公式として不適切なものを使用すると、確率波高の推定値に偏りを生じる。本稿で提示した式(3)~(7)の諸公式は、母集団の分布関数を適正に見定めているかぎり、偏りを生じるおそれは少ない。しかし、図-5 で例示したように、極値データのサンプルに特有な統計的変動性によって母関数以外の分布関数があてはめられてしまう場合がある。その場合には、確率波高の推定値が母集団の値からずれる結果となる。図-7 はこの一例である。

ここでは、本稿で提示している5種類の極値分布関数の各々から、データ個数 $N=40$ のサンプルを一様乱数を使って10,000組抽出した。全体で50,000組である。データは全て使用したので全数極値資料 ($\nu=1$) としての取り扱いである。各サンプルに分布関数のあてはめを行なった結果から、 $k=0.75$ のワイブル分布が最適と判定されたのは8,722組であり、母分布関数ごとの組数は図中に

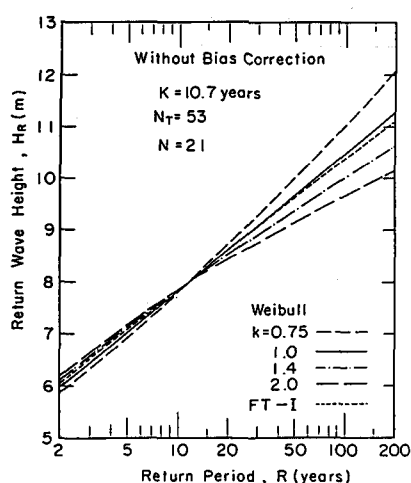


図-6 各種分布関数に基づく確率波高の推定結果

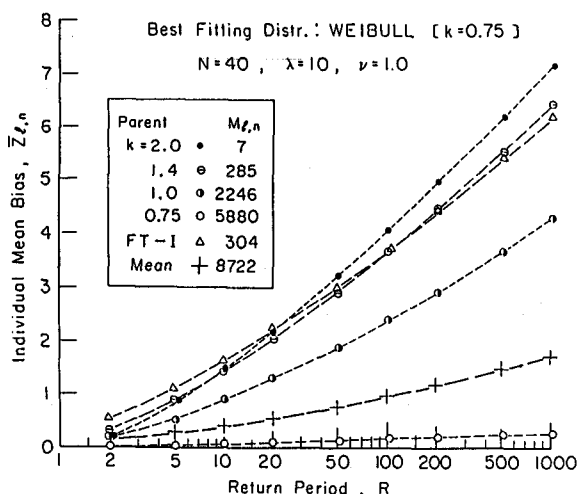


図-7 母分布関数が未知であることによる確率波高の推定値の偏り量 $Z_{L,n}$ の算出例

$M_{l,n}$ として記入してある。たとえば、FT-I型分布10,000組のうち304組が $k=0.75$ のワイブル分布に帰属すると見誤られている。これらのサンプルについて確率波高を計算すると、母関数であるFT-I型ではなく $k=0.75$ のワイブル分布で推定されるため、真の値よりも大きめの値が得られる。図の縦軸はこの偏り量を無次元化して表示している。無次元化は次式による。

$$\bar{Z}_{l,n} = E[(\hat{x}_{R,n} - x_{R,l})/\sigma_x] \quad (22)$$

ここに、 $E[\]$ はアンサンプル平均、 $\hat{x}_{R,n}$ は n 番目の分布関数をあてはめて求めた R 年確率波高の推定値、 $x_{R,l}$ は母集団である l 番目の分布関数における真の確率波高、 σ_x はサンプルの標準偏差である。すなわち、推定値の平均が真の値から標準偏差の何倍離れているかを表している。

こうした数値実験では母関数が最初から与えられているけれども、実際の極値統計解析では母分布関数が未知である。また、ここで取り上げた5種類の極値分布がどれだけの頻度で出現するかも明らかでない。しかし、何らかの補正を行なう必要性はある。そこで、この5種類の極値分布が等しい確率で自然界に存在するものと仮定して偏り量 $\bar{Z}_{l,n}$ の重み付き平均を用いることにする。図中の+印はその結果である。

(3) 確率波高の偏り補正の実験式

図-7に示したような数値実験は、サンプル中のデータ個数が $N=40$ の場合だけでなく、 $N=10, 20, 100$, および200の場合についても実施し、各サンプルに対して最適と判定された分布関数の各々について偏り量の平均値 $\bar{Z}_n$ を計算した。図-8は $k=1.4$ のワイブル分布が最適とあてはめられた場合の $\bar{Z}_n$ の結果である。このあてはめ関数の場合、母関数が $k=0.75$ や1.0などであると、確率波高の推定値は真の値よりも小さく出る。したがって、 $\bar{Z}_n$ は全体として負の値となる。

数値実験は全数極値資料だけでなく、 $\nu=0.50$ および0.25の部分極値資料についても行なった。そして、これらの結果から $\bar{Z}_n$ に関する実験式を次のように取りまとめた。

$$\bar{Z}_n = \begin{cases} A_c (y_R + \alpha \ln \nu)^q & : y_R > -\alpha \ln \nu \\ 0 & : y_R \leq -\alpha \ln \nu \end{cases} \quad (23)$$

ここに、係数 A_c と α , および指数 q は分布関数毎にそれぞれ次のように定めた。

1) ワイブル分布 ($k=0.75$)

$$A_c = \begin{cases} 0.030 \exp[-0.6(\log_{10} N/4)^2] & : \nu=1.0 \\ 0.025 \exp[-0.7(\log_{10} N/15)^2] & : \nu=0.50, 0.25 \end{cases} \quad (24)$$

$$q=1.6, \quad \alpha=2.7$$

2) ワイブル分布 ($k=1.0$)

$$A_c = \begin{cases} 0.028 N^{-0.25} & : \nu=1.0 \\ -0.0028 - 0.006(\log_{10} N/50)^2 & : \nu=0.50, 0.25 \end{cases} \quad (25)$$

$$q=2.1, \quad \alpha=1.0$$

3) ワイブル分布 ($k=1.4$)

$$A_c = \begin{cases} -0.40 N^{-0.8} & : \nu=1.0 \\ -0.10 N^{-0.4} & : \nu=0.50, 0.25 \end{cases} \quad (26)$$

$q=2.7, \alpha=0.5$

4) ワイブル分布 ($k=2.0$)

$$A_c = \begin{cases} -0.50 N^{-0.7} & : \nu=1.0 \\ -0.64 N^{-0.6} & : \nu=0.50, 0.25 \end{cases} \quad (27)$$

$q=3.4, \alpha=0.35$

5) FT-I型分布

i) $\nu=1.0$

$$A_c = \begin{cases} 0.046 - 0.40 [\log_{10} (60/N)]^3 & : N < 60 \\ 0.046 \exp \{-2.5 [\log_{10} (N/60)]^2\} & : N \geq 60 \end{cases} \quad (28)$$

$q=1.0, \alpha=0.9$

ii) $\nu=0.50, 0.25$

$$A_c = 0.01 - 0.044 [\log_{10} (N/300)]^4 \quad (29)$$

$q=1.0, \alpha=0.9$

図-8の中の曲線は上記の実験式によるあてはめ結果である。

データ採択率が $\nu=1.0, 0.5$, および 0.25 以外については未検討である。ただし、これらの実験式はある程度のあてはめ誤差を許容して作成したものであるため、当分の間は $\nu > 0.8$ のデータに対しては $\nu=1.0$ の実験式を適用し、 $\nu \leq 0.8$ に対しては $\nu=0.50, 0.25$ の諸式を使うのがよいのではないかとと思われる。ただし、データ採択率のあまり低いところまで適用するのは補正過多となるおそれがあるので、 $\nu=0.15$ あたりを下限值としておくのがよいであろう。

(4) 確率波高の偏りの補正方法

確率波高の推定値の偏りを補正する方法としては、今回の数値実験で用いた式(2)の無次元化の形を利用し、次のようにすることが考えられる。

$$(x_R)_{cor} = \hat{x}_{R,n} - \bar{Z}_n \sigma_x \quad (30)$$

ただし、この補正方法にしても便宜的なものにすぎない。というのは、この式の中の σ_x が与えられた極値データの標準偏差であるため、そのサンプルの σ_x が大きければ補正量も大きくなり、 σ_x が小さければ補正量も小さくなる

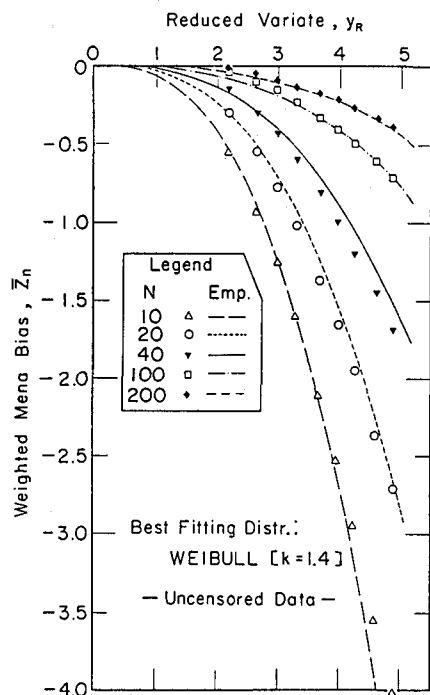


図-8 最適あてはめ関数による
確率波高の偏り量の平均値
 $\bar{Z}_n$ の算出例

ためである。すなわち、極値統計データの変動性の影響は確率波高の補正に関しても現われることになる。

補正方法を具体的に説明するため、表-1のデータを使って計算例を示す。このデータに対して最も適合すると判定された $k=2.0$ のワイブル分布を使って50年確率波高を求め、偏り補正を行なう。まず式(21)で基準化変量を計算すると、

$$y_{50} = [\ln(4.93 \times 50)]^{1/2} = 2.347$$

したがって、50年確率波高は表-1に記載した母数推定値 $\hat{A}$ および $\hat{B}$ を使って次のように求められる。

$$H_{50} = 3.560 \times 2.347 + 0.786 = 9.14 \text{ m}$$

これに対する偏り補正量を求めるため式(27)で係数 A_c を計算すると、 $\nu=0.396$ であるから

$$A_c = -0.64 \times 21^{-0.6} = -0.103$$

したがって、

$$\begin{aligned} \bar{Z}_n &= -0.103 \times (2.347 + 0.35 \times \ln 0.396)^{3.4} \\ &= -1.13 \end{aligned}$$

このデータの標準偏差は $\sigma_x = 1.101 \text{ m}$ であるので、偏り補正量は

$$\Delta x_{50} = \bar{Z}_n \sigma_x = -1.13 \times 1.101 = -1.24 \text{ m}$$

この結果、50年確率波高は次のように推定される。

$$x_{50} = 9.14 - (-1.24) = 10.38 \text{ m}$$

このようにして表-1のデータに対して確率波高を推定し、偏り量を補正した結果を表-2に示す。

この偏り補正の結果によると、各分布関数のあてはめによる確率波高の大小の順位が逆転し、 $k=2.0$ のワイブル分布が最大の推定値を与えるようになっている。なお、この表には次節で述べる標準誤差の値も載せてある。

6. 確率波高の信頼区間の計算

既に図-1で例示したように、極値データとして与えられたサンプルはかなり大きな変動性を伴っている。したがって、与えられた資料に対して確率波高を計算し、偏り補正を行なったとしても、その結果は必ずしも母集団の真の値に一致せず、真値のまわりに統計的に分散する。この分散の大きさを標準誤差の形で表し、前節で述べた数値実験のデータを使って実験式としてまとめたのが以下である。

$$\begin{aligned} S_n &= \{E[(\hat{x}_{R,n} - x_R)^2 / \sigma_x^2]\}^{1/2} \\ &= \{1.0 + A_s | y_R + \alpha \ln \nu |^p\} / \sqrt{N} \end{aligned} \quad (31)$$

ここに、 $\hat{x}_{R,n}$ は n 番目の分布関数をあてはめて算定した R 年確率波高の推定値、 x_R は真値、 σ_x は極値データのサンプルの標準偏差であり、係数 A_s および指数 p は次式のように設定されている。なお、

係数 α は偏り量に対する補正の式(24)~(29)と共通である。

1) ワイブル分布 ($k=0.75$)

$$A_s = \begin{cases} 0.57 + 0.18 [\log_{10} (N/20)]^2 : \nu=1.0 \\ 0.41 + 0.22 [\log_{10} (N/20)]^2 : \nu=0.50, 0.25 \end{cases} \quad (32)$$

$p=1.2, \alpha=2.7$

2) ワイブル分布 ($k=1.0$)

$$A_s = \begin{cases} 0.55 + 0.15 [\log_{10} (N/15)]^2 : \nu=1.0 \\ 0.38 + 0.17 [\log_{10} (N/20)]^2 : \nu=0.50, 0.25 \end{cases} \quad (33)$$

$p=1.7, \alpha=1.0$

3) ワイブル分布 ($k=1.4$)

$$A_s = \begin{cases} 0.37 + 0.08 [\log_{10} (N/1000)]^2 : \nu=1.0 \\ 0.46 + 0.09 [\log_{10} (N/20)]^2 : \nu=0.50, 0.25 \end{cases} \quad (34)$$

$p=2.3, \alpha=0.5$

4) ワイブル分布 ($k=2.0$)

$$A_s = \begin{cases} 0.30 + 0.36 [\log_{10} (N/80)]^2 : \nu=1.0 \\ 0.56 + 0.20 [\log_{10} (N/100)]^2 : \nu=0.50, 0.25 \end{cases} \quad (35)$$

$p=3.2, \alpha=0.35$

5) FT-I型分布

$$A_s = \begin{cases} 0.24 + 0.36 [\log_{10} (N/80)]^2 : \nu=1.0 \\ 0.46 + 0.14 [\log_{10} (N/50)]^2 : \nu=0.50, 0.25 \end{cases} \quad (36)$$

$p=1.6, \alpha=0.9$

データ採択率が1.0, 0.5, および0.25以外については、確率波高の偏り補正の実験式と同様に取り扱うものとする。

以上の実験式の使用例として、前節で求めた50年確率波高に対する標準誤差を計算する。あてはめ関数は $k=2.0$ のワイブル分布を使うことにして係数 A_s を求めると、

$$A_s = 0.56 + 0.20 \times [\log_{10} (21/100)]^2 = 0.652$$

したがって、標準誤差の無次元量 S_n は

$$S_n = [1.0 + 0.652 \times |2.347 + 0.35 \times \ln 0.396|^{3.2}] / \sqrt{21} \\ = 1.57$$

標準誤差の絶対値は、式(31)の形を利用して次のように計算すればよいと考えられる。

$$\sigma_R = S_n \sigma_x \quad (37)$$

上の例であれば、

$$\sigma_{50} = 1.57 \times 1.101 = 1.73 \text{ m}$$

表－２ 確率波高とその標準誤差の計算結果

再現期 間(年)	FT－Ⅰ型分布		ワイブル ($k=0.75$)		ワイブル ($k=1.0$)		ワイブル ($k=1.4$)		ワイブル ($k=2.0$)	
	H_R	$\sigma(H_R)$	H_R	$\sigma(H_R)$	H_R	$\sigma(H_R)$	H_R	$\sigma(H_R)$	H_R	$\sigma(H_R)$
2.0	6.16(−0.11)	0.44	5.87(+0.01)	0.28	6.01(−0.01)	0.39	6.17(−0.07)	0.46	6.38(−0.20)	0.51
5.0	7.27(−0.18)	0.69	6.83(+0.01)	0.50	7.09(−0.04)	0.61	7.29(−0.17)	0.69	7.58(−0.42)	0.77
10.0	8.09(−0.23)	0.93	7.58(+0.21)	0.70	7.92(−0.07)	0.82	8.11(−0.27)	0.90	8.44(−0.62)	1.02
20.0	8.90(−0.28)	1.20	8.34(+0.37)	0.94	8.75(−0.11)	1.07	8.92(−0.39)	1.16	9.28(−0.87)	1.30
50.0	9.98(−0.35)	1.60	9.36(+0.64)	1.30	9.87(−0.18)	1.45	9.97(−0.59)	1.54	10.38(−1.24)	1.73
100.0	10.79(−0.40)	1.94	10.13(+0.89)	1.60	10.73(−0.24)	1.78	10.77(−0.76)	1.87	11.22(−1.57)	2.10

注：括弧内は偏り補正量であり，確率波高は補正後の値である。

表－２に記載した標準誤差は以上のようにして計算したものである。

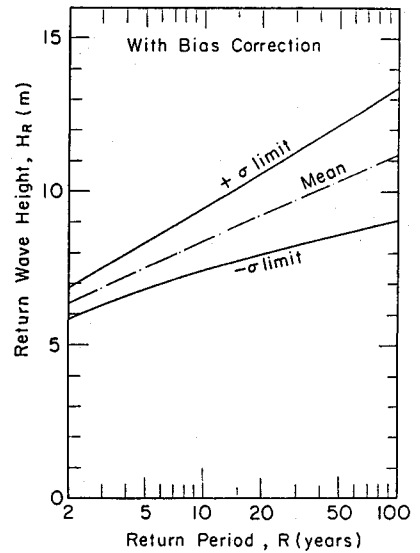
確率波高の推定値の統計的分布の関数形は，データ個数やあてはめ関数の特性によって変化し，明確に表すことがむずかしい。しかし，信頼区間の推定としては近似的に正規分布を用いることにすると，上で求めた標準誤差の値が信頼区間の幅を表す指標となる。したがって，１シグマ限界を採用すると，上述の50年確率波高は

$$H_{50} = (10.38 - 1.73) \sim (10.38 + 1.73) \\ = 8.6 \sim 10.5 \text{ m}$$

とある幅の中で推定されることになる。２シグマ限界をとると， $H_{50} = 6.9 \sim 13.8 \text{ m}$ となる。図－９は１シグマ限界を用いた場合の確率波高の分布範囲を示したものである。当然のことながら，再現期間の長いところについては，確率波高の推定誤差も大きくなる。

従来の確率波高の計算では単一の値を算出してはいたけれども，これは極値データの統計的誤差を無視していたものである。本来は，この例のように信頼区間の幅を付記して表示すべきである。こうした信頼区間の幅の中のどのような値を設計波高として採択するかは主任技術者の重要な判断事項である。使用した波浪の極値データの信頼度，設計対象の構造物の重要度その他を勘案して決めなければならない。同一地区の構造物であっても，設計波高が異なることがありうる。確率波高の信頼区間を考慮して設計波高を選定する方法については，今後いろいろ検討していく必要がある。

一方，極値統計解析としては信頼区間の幅をできるだけ狭める工夫しなければならない。このためには，式(31)の形から分かるように，データ個数 N を増す必要があり，解析対象の統計期間をできるだけ長く取るとともに，波高の低いところまで洩れなく拾い上げる努力が望まれる。



図－９ 確率波高の推定値の信頼区間の例

7. 気象原因別の極値資料の取り扱い

極値統計の解析を行なう場合、取り扱うデータが同じ母集団から抽出されたものであることが必要であり、これは統計学の基本原則の一つとされている。同一の母集団に属することの証明は困難であるので、対象とする現象の発生メカニズムを考えて判断する。たとえば、台風による強風と高気圧からの吹き出しによる強風とでは風速の絶対値に差があることは容易に推察されるので、これらは別個に解析すべきである。波浪の場合、台風による高波と低気圧あるいは冬期季節風による高波とを区別すべきかどうかは、ややむずかしい問題である。波高の絶対値に差がある場合には分けて処理すべきであるが、絶対値の差がほとんどない場合などは、個別解析と総合解析とを試み、その結果を見て判断する必要がある。

高波の極値データを気象原因別に分けて解析した場合、その結果をどのようにまとめるかという問題が生じる。これに関しては、Carter & Challenor¹¹⁾の研究が一つの示唆を与える。この研究は、風速や波高のように年間で季節的に変化する現象は月毎の統計分布が異なる、すなわち母集団が異なることを指摘し、単純に毎年最大値資料を使って解析すると、結果として確率風速や確率波高を低めに推定することを例示したものである。この解決策として二人は毎月最大値資料を使って月毎の極値分布を求め、月毎の非超過確率の乗積としての全体の極値分布を計算することを提案した。

この考え方は、複数の気象原因別の高波の統計資料にも適用することができる。ただし、本稿で扱っているような極大値資料の場合には、気象原因毎の高波の平均発生率が異なるので、この調整が必要である。具体的には、極大値の年間発生回数がポアソン分布で近似できると仮定し、極大値に関する分布関数 $F(x)$ を毎年最大値に関する分布関数 $F_*(x)$ に変換する。この結果は

$$F_*(x) = \exp \{-\lambda [1 - F(x)]\} \quad (38)$$

となる¹⁾。いま、 n 個の気象原因別の極値分布関数が求められたとすると、全体としての分布関数は気象原因別の極値分布関数の乗積として次式のように計算される。

$$\begin{aligned} F(x) &= \prod_{i=1}^n F_{*i}(x) \\ &= \exp \left\{ \sum_{i=1}^n \lambda_i [1 - F_i(x)] \right\} \end{aligned} \quad (39)$$

ここに示した方法の問題点の一つは、母分布関数が未知であることに起因する確率波高の偏りの補正がむずかしいことである。各原因別の極値データについて偏り補正の大きさを計算しておき、この結果を参照して全体の確率波高に対する補正量を目分量で決めるというのが当面の方法であろう。

全体としての確率波高の標準誤差については、次式を用いて推定するのが一つの方法である¹⁾。

$$\sigma_{\text{all}}(\hat{x}_R)^2 = \left\{ \sum_{i=1}^n N_i [1 - F_i(\hat{x}_R)] \times \sigma_i(\hat{x}_R)^2 \right\} / \sum_{i=1}^n N_i [1 - F_i(\hat{x}_R)] \quad (40)$$

ただし、 N_i は i 番目の発生原因別資料におけるデータ個数である。実際の計算においては、全体としての乗積確率として求めた分布関数 $F(x)$ と x との数値計算結果に基づき、内挿によって所定の再現期間 R に対する確率波高 $\hat{x}_R$ をまず求める。そして、その基礎データである個々の分布関数 $F_i(x)$ について同じく内挿によって $F_i(\hat{x}_R)$ を求めればよい。発生原因別のデータの標準誤差

$\sigma_i(\hat{x}_R)$ は式(31)によって基準化変量 y_R の関数として直接に求められる。

8. あとがき

極値統計は常に不確定性の問題を伴っている。現在までに得られた資料に基づいて確率波高を推定しても、次年度以降にどのような高波が来襲するかを予測することはできない。基本的には、われわれが手にしている極値データが一つの統計的サンプルであり、確率のいたずらによって母集団の平均値よりも大きなものを得ているのかもしれないし、その逆であるかもしれず、しかもそれを知る手段がないことである。どのような方法で確率波高を推定しても、それは確定値ではあり得ない。必ず統計的誤差を伴っている。本稿で紹介したのは、この誤差の大きさを具体的に推定する方法である。さらに、ここでは取り上げなかったが、自然現象が数十年以上の規模で変化する長期変動の問題も残されている。

最近信頼性設計の手法が脚光を浴びており、設計対象の外力の変動性を考慮できるようになっている。しかし、この場合でも極値分布関数はデータにあてはめて算定したものを使うので、それが母集団の分布関数とずれている可能性の問題は依然として残されている。

横造物を設計する技術者として、設計条件が不確定であるのは不安なことである。しかし、本来自然を相手とする以上、こうした不確定性はやむを得ないところである。今までは、極値統計解析を行なうと、たとえば100年確率波高が確定値として得られるような錯覚が一部にあったのではないかと思われる。このような誤解は速やかに取り去り、技術者としての健全な判断を活かすようにしていく必要がある。本稿がそうした方向へ進むために役立てば何よりも幸いである。

参 考 文 献

- 1) 合田良実：極値統計におけるプロット公式ならびに推定値の信頼区間に関する数値的検討，港湾技術研究所報告，第27巻，第3号，1988. 3， pp. 31-92.
- 2) 神田 徹・藤田睦博：「水文学-確率論的手法とその応用」，技報堂出版，1982，275 p.
- 3) 合田良実：波浪統計に関する二，三の考察，港湾技研資料，No. 39，1976， pp. 239-255.
- 4) Gumbel, E. J. : *Statistics of Extremes*, Columbia Univ. Pres., New York, 1958. (河田・岩井・加瀬監訳：「極値統計学」，生産技術センター新社)
- 5) Muir, L. R. and El-Shaarawi, A. H. : On the calculation of extreme wave heights : a review, *Ocean Engng.*, Vol. 13, No. 1, 1986, pp. 93 - 118.
- 6) Carter, D. J. T. and Challenor, P. G. : Methods of fitting the Fisher-Tippett type I extreme value distribution, *Ocean Engng.*, Vol. 10, No. 3, 1983, pp. 191 - 199.
- 7) Cunnane, C. : Unbiased plotting positions—a review, *J. Hydrology*, Vol. 37, 1978, pp. 205 - 222.

- 8) Blom, G. : *Statistical and Transformed Beta-Variates*, John Wiley & Sons, New York, 1958, Chapt. 12.
- 9) Gringorten, I. I. : A plotting rule for extreme probability paper, *J. Geophys. Res.*, Vol. 68, No. 3, 1963, pp. 813 — 814.
- 10) Petruaskas, C. and Aagaard, P. M. : Extrapolation of historical storm data for estimating design wave heights, *Prepr. 2nd Ann. Offshore Tech. Conf.*, 1970, OTC1190.
- 11) Carter, D. J. T. and Challenor, P. G. : Estimating return values of environmental parameters, *Quart. J. Roy Met. Soc.*, Vol. 107, 1981, pp. 259 — 266.